Journal of Research on Educational Effectiveness, 8: 218–244, 2015

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2014.909548

Routledge
Taylor & Francis Group

Using Regression Discontinuity to Test the Impact of a Tier 2 Reading Intervention in First Grade

Scott K. Baker
University of Oregon, Eugene, Oregon, USA
Southern Methodist University, Dallas, Texas, USA

Keith Smolkowski
Oregon Research Institute, Eugene, Oregon, USA

Erin A. Chaparro, Jean L. M. Smith, and Hank Fien
University of Oregon, Eugene, Oregon, USA

Abstract: Multitiered systems of reading instruction and intervention, including response to intervention, are widely used in early reading by schools to provide more intense services to students who need them. Research using randomized controlled trials has compared innovative Tier 2 interventions to business-as-usual Tier 2 approaches and established a number of important components that compose effective Tier 2 interventions in early reading. The purpose of this study was to test the impact of a Tier 2 intervention with Tier 2 compared to Tier 1 instruction alone using regression discontinuity. A cut score was used to assign first-grade students at risk for reading difficulties to Tier 2 intervention plus Tier 1 instruction. Students who missed the cut score in the control group received Tier 1 instruction only. Students in the treatment group, just below the cut score, made greater gains on the SAT10 total score and the individual subtests than students just above the cut score. Outcomes were not significant on oral reading fluency.

Keywords: Regression discontinuity design, early reading instruction, tiered instruction and intervention, reading risk, responsibility to intervention

The reauthorization of the Individuals with Disabilities Improvement Act (IDEA; 2004) has led to the increased use of school-wide, multitiered approaches of instructional supports and services for students (Burns & Gibbons, 2012; Denton, 2012). The most common tiered approach is the three-tier model (Chard et al., 2008; Denton, 2012; Haager, Klingner, & Vaughn, 2007a, 2007b; Spectrum K12 School Solutions, 2010; Walker & Shinn, 2002). Tier 1 instruction represents the core program and instruction is differentiated for students to increase their chances of reaching important learning goals at designated time points (Connor et al., 2011; Fuchs, Fuchs, & Stecker, 2010; McMaster, Kung, Han, & Cao, 2008). Students who struggle to make expected progress in Tier 1, even with differentiated instruction, receive Tier 2 intervention. The purpose of Tier 2 is to provide intervention

Address correspondence to Scott K. Baker, University of Oregon, Center on Teaching and Learning, 1600 Millrace Drive, Suite 207, Eugene, OR 97403, USA. E-mail: sbaker@uoregon.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uree.

that is supplemental to Tier 1, so as to accelerate students' progress toward important learning goals. In beginning reading, the most common structure of Tier 2 intervention is 20 to 40 min of daily instruction delivered in small-group formats (Gersten et al., 2009). Students whose response to Tier 2 intervention does not result in expected rates of progress are provided with Tier 3 intervention, which is highly intense and can include a combination of small group and one-on-one intervention.

The intervention investigated in this study represents a systems-level approach to literacy instruction called Enhanced Core Reading Instruction (ECRI). The ECRI system includes three main parts: (a) enhanced core instruction in Tier 1 to make it more accessible to students at risk for reading difficulties, (b) a Tier 2 intervention that is closely aligned with Tier 1 in terms of instructional content and delivery formats, and (c) using multiple sources of formative data to make instructional decisions. The current study uses regression discontinuity (RD) to examine whether the ECRI Tier 2 intervention was effective for students at risk for reading difficulties who received Tier 1 instruction plus Tier 2 intervention. The performance of these students was compared to similar students who received Tier 1 instruction only.

RESEARCH BASE ON TIER 2 READING INTERVENTIONS

The Institute of Education Sciences released a practice guide on tiered RTI approaches in the primary grades, and one recommendation—small-group intervention in Tier 2—was rated as having the strongest level evidence to support the practice (Gersten et al., 2009). Multiple studies using randomized controlled trials (RCTs) contributed to this recommendation and the panel concluded that approximately 30 min of small-group intervention each day (3–5 days per week) is highly effective for students who are struggling with learning to read.

In the development of the ECRI system we followed the instructional features that were common across the practice guide studies addressing how teachers delivered instruction during the small-group intervention. Nearly all recommendations of small-group instruction in beginning reading, based on both the RTI practice guide (Gersten et al., 2009) and other studies and sources (Connor et al., 2009; Elbaum, Vaughn, Hughes, & Moody, 1999, 2000; Evans, 1996; Gunn, Smolkowski, & Vadasy, 2011; Kirschner, Sweller, & Clark, 2006; Taylor, Pearson, Clark, & Walpole, 1999; Thurlow, Ysseldyke, Wotruba, & Algozzine, 1993; Vaughn, Linan-Thompson, Kouzekanani, et al., 2003), highlight the underlying importance of providing explicit instruction. Explicit instruction in reading includes these key features: (a) teachers model, demonstrate, and clearly explain the skills and knowledge they expect students to apply (Mathes et al., 2005); (b) teachers provide multiple opportunities for students to practice learning objectives during the lesson (Gunn, Smolkowski, Biglan, Black, & Blair, 2005; Smolkowski & Gunn, 2012; Vadasy, Sanders, & Peyton, 2005); (c) teachers correct and attend to student errors and misunderstandings immediately and systematically (Gunn et al., 2005); and (d) teachers pace lessons briskly so that student engagement is high and multiple areas of reading development (e.g., phonemic awareness, vocabulary, word reading) can be addressed during the lesson (Sutherland, Alder, & Gunter, 2003).

EVALUATING TIER 2 IMPACT IN TIERED SETTINGS

One important way of evaluating the impact of a reading intervention is to compare outcomes from the implementation of two similar interventions. All 11 Tier 2 studies in the RTI

Practice Guide used RCTs to evaluate whether innovative Tier 2 interventions (treatment condition) outperformed the standard Tier 2 interventions being used (control condition, typically business as usual). A second way of demonstrating impact is to compare Tier 2 interventions with Tier 1 instruction.

Evidence that a Tier 2 intervention works better than other Tier 2 approaches does not tell us what the impact of the intervention is when it is compared with Tier 1 instruction. This knowledge is important because the vast majority of students who receive a Tier 2 reading intervention also receive the bulk of their reading instruction in Tier 1, alongside peers who only receive Tier 1 instruction (Hill, King, Lemons, & Partanem, 2012; Wanzek & Cavanaugh, 2012). It would be useful to have evidence of increased growth of students assigned to receive Tier 2 intervention compared to those who receive only Tier 1 because different growth patterns should lead to the continued success of Tier 2 students when the Tier 2 supports are removed (Rodden-Nord, Shin, & Good, 1992; Shinn, Baker, Habedank, & Good, 1993; Shinn, Habedank, & Baker, 1993). Gunn et al. (2005), for example, showed continued increased growth of treatment students compared to controls in oral reading fluency 2 years after the end of small-group intervention.

Students at risk for reading difficulties who acquire skills commensurate with typically achieving students through intervention has implications for their likely success in Tier 1 when intervention supports (e.g., Tier 2) are removed (Rodden-Nord et al., 1992; Shinn, Baker, et al., 1993; Shinn, Habedank, et al., 1993). Effective Tier 2 interventions should be able to demonstrate that students who receive Tier 2 intervention (a) experience accelerated growth as compared to similarly skilled students who receive a different Tier 2 intervention and (b) experience accelerated growth as compared to peers demonstrating grade level skills in Tier 1 instruction. Studies in each area provide independent estimates of scientific evidence regarding the efficacy of the intervention.

The rationale for comparing student performance in Tier 2 with student performance in Tier 1 is also based on the assumption that the majority of students who receive Tier 2 intervention also receive Tier 1 instruction (Hill et al., 2012; Wanzek & Cavanaugh, 2012). Thus, a Tier 2 intervention that demonstrates larger than expected gains for students who receive Tier 1 instruction and Tier 2 intervention compared to students who receive Tier 1 instruction only provides evidence that students who receive both Tier 1 and Tier 2 are developing the knowledge and skills necessary to perform on par with their peers who receive Tier 1 only. This evidence would support the position that many students at risk for reading difficulties may be able to continue to learn successfully with only Tier 1 instruction after they have completed the Tier 2 intervention (Rodden-Nord et al., 1992; Shinn, Baker, et al., 1993; Shinn, Habedank, et al., 1993).

STUDY CONTEXT

In a randomized trial, Fien et al. (2014) compared the performance of students at risk for reading difficulties in schools where they received the ECRI Tier 2 treatment to the performance of students at risk for reading difficulties in schools where they received the district's standard Tier 2 intervention. Smith et al. found statistically significant differences for students who received Tier 2 in 22 ECRI schools, compared to similar students in 22 control schools, on fall-to-winter gains in decoding and oral reading fluency (effect sizes from .21 to .30) and fall-to-spring gains in decoding and word reading (effect sizes from .24 to .32). The current study extends this work to compare the performance of students who

received the ECRI Tier 2 treatment (ECRI Tier 1 instruction *and* ECRI Tier 2 intervention) to students who received only ECRI Tier 1 instruction.

The present study uses an RD design, where students who scored at or below the 30th percentile on the Stanford Achievement Test Tenth Edition (SAT10; Harcourt Educational Measurement, 2002) were assigned to receive a Tier 2 intervention in addition to Tier 1 instruction (intervention students). Students above the 30th percentile received only Tier 1 instruction (control students). As is standard in RD, the performance of students close to the cut score is the main focus of analysis and conclusions are drawn about what would have occurred with control students close to the cut score had they received the intervention (Smolkowski, Strycker, & Seeley, 2013). We could find no previous studies that used a rigorous design to evaluate the impact of a Tier 2 intervention in relation to Tier 1 instruction.

With this RD design, we address one primary research question: Do students who receive both ECRI Tier 1 instruction and ECRI Tier 2 intervention outperform similar students who receive ECRI Tier 1 instruction only?

METHOD

The analysis sample included students from 22 ECRI schools. Each school participated in the study for 1 year; eight schools began in 2009 and 14 began in 2010. Schools implemented ECRI for Tier 1, and the ECRI Tier 2 intervention was implemented for students who met criteria, which was based on a cut-off score from the SAT10 (30th percentile). Consistent with the RD design, the Tier 1 students served as the controls for the Tier 2 treatment students.

These schools also participated in an RCT (Fien et al., 2014; Smith et al., 2013) that compared ECRI Tier 2 students in the 22 schools to similar Tier 2 students in 22 control schools. In this RD study, we used the 22 control schools from the RCT to estimate the functional form of the relationship between pretest and posttest scores on dependent variables and to examine the discontinuity in the regression line at the cut point with students who did not receive the ECRI Tier 2 supports (Wing & Cook, 2013). The RD design elements (e.g., identification processes, cut scores for assignment to Tier 2) were identical in both conditions. The 22 control schools are described more fully in Fien et al. (2014) and Smith et al. (2013).

Recruitment and Sampling Procedures

District Recruitment

We asked districts in two regions of the country (western Oregon and eastern Massachusetts) to help us identify schools that might be interested in participating in the ECRI project. We recruited districts with schools that had three features of RTI in place. First, the district used a formative assessment system in all of their schools. This characteristic demonstrated for us that any participating school would have an important feature of RTI in place. Second, all schools in the district used a published, core program for Tier 1 reading instruction. These criteria were necessary to ensure that we would be able to develop the Tier 2 intervention. Third, all schools had a system in place for providing reading intervention for students at risk for reading difficulties (e.g., Tier 2 intervention). This established system ensured that

Table 1. Student-level and school-level demographic information

	Tier 1		Tier 2
Student level			
Students identified as eligible for limited English proficient	9%		21%
Students identified as eligible for special education	3%		8%
		All Tiers *M* (*SD*)	
School level			
Students eligible for free or reduced lunch program		49% (.20)	
White, non-Hispanic		62% (.22)	
Hispanic or Latino		20% (.14)	
Asian or Pacific Islander		08% (.09)	
Black or African Americans		04% (.07)	
American Indian		01% (.01)	
Hawaiian or Pacific Islander		01% (.02)	
Multiracial		02% (.03)	

schools assigned to the ECRI condition would be able to implement the treatment Tier 2 intervention and that schools assigned to the control condition would have a system in place for providing business as usual Tier 2 intervention to students. We approached 10 school districts that met these criteria, and nine chose to participate—six from Oregon and three from Massachusetts. These districts then helped us recruit schools that would agree to the study requirements.

School Recruitment

Participant schools agreed to (a) 90 min of core reading instruction for all students, and (b) 30 min of daily small-group intervention for students at risk for reading difficulty. In Year 1, 18 schools agreed to participate. These schools were randomly assigned to the ECRI or control condition while blocking on district. Two schools—one treatment and one control—withdrew shortly after randomization, leaving 16 schools that completed the study. The eight ECRI schools compose the Cohort 1 sample for the RD analysis. In Year 2, 28 additional schools agreed to participate. Blocking on district, these schools were randomly assigned to the ECRI or control condition (14 in each condition). Thus, across the 2 years, 22 schools implemented the ECRI intervention and are included in RD analysis. The 22 control schools implemented their standard approach for Tier 1 instruction and Tier 2 intervention and compose the secondary sample. Table 1 displays school level race, ethnicity, and eligibility for free and reduced lunch data for the 22 ECRI intervention schools.

Student Recruitment

After schools were assigned to condition, letters requesting (passive) consent were sent to parents and legal guardians of first-grade students. Students of parents who declined

participation were not administered study assessments and the school determined which reading services and supports these students received. The percentage of students for whom parents declined was less than 1%. No students withdrew from the study during first grade.

Participants

Students

In the 22 ECRI schools, 1,644 students attended first-grade classrooms. Students with consent were administered the SAT10 at pretest and a cut score was used to determine the type of reading service and support students received ($n = 1{,}509$). A total of 135 students did not have pretest scores because of incomplete pretest data or testing occurred prior to the student's entry into the school. Students who scored above the 30th percentile received ECRI Tier 1 instruction ($n = 819$). Students who scored at or below the 30th percentile received ECRI Tier 2 reading intervention. Intervention students who scored at or above the 10th percentile received only ECRI Tier 2 intervention (in addition to ECRI Tier 1 instruction; $n = 392$). Students who scored below the 10th percentile may have also received more intensive intervention services designed by the district and schools ($n = 298$). Table 1 provides the percentage of students identified as eligible for limited English proficiency services and special education services at the student level for students in ECRI Tiers 1 and 2. Students in control schools were identified for reading services using the same procedures as students in ECRI schools.

Teachers

Certified classroom teachers provided Tier 1 instruction. A total of 69 first-grade teachers, with an average of 12.2 years of teaching experience, were trained and provided ECRI Tier 1 instruction. These teachers were paid a $500 stipend for participation.

ECRI Interventionists

To increase intervention fidelity, ECRI recruited and paid 67 interventionists to deliver the Tier 2 intervention. Interventionists were trained by, supervised by, and reported regularly to ECRI research staff. Interventionists included five certified teachers and 62 instructional assistants who, in most cases, were also employed by the schools in which they provided the Tier 2 intervention. The majority of the interventionists were female ($n = 65$; 97%) and approximately evenly divided between those with 5 years or less of experience in educational settings ($n = 35$; 52%), and those with more than 5 years of experience ($n = 32$; 48%). Most had previous experience teaching small groups ($n = 41$; 61%).

ECRI Intervention System

Three features distinguish the ECRI intervention system from other tiered systems of instructional support. The first feature is a redesigned core program following principles of instructional design to increase the explicitness of Tier 1 instruction (Baker, Fien, & Baker, 2010; Coyne, Kame'enui, & Carnine, 2011; Harn, Kame'enui, & Simmons, 2007; National Association of State Directors of Special Education, 2006; Vaughn, Linan-Thompson, &

Hickman, 2003; Vellutino et al., 1996). The second is increased instructional specificity in Tier 1, achieved through daily *lesson maps* and *instructional routines* that teachers use to address lesson objectives. The third is close alignment between Tier 1 instruction and Tier 2 intervention. The Tier 2 intervention was delivered to groups of approximately five students and used the same scope and sequence of skills and highly aligned content that was used during Tier 1 instruction.

Tier 1 Enhancements Overview

Table 2 juxtaposes the enhancements as they are applied to Tier 1 instruction and Tier 2 intervention. The cornerstone of the ECRI Tier 1 instructional enhancements was the use of two instructional tools. First, daily lesson maps accompanied each core reading program (Ashlock Consulting, 2006; Kame'enui, Simmons, Coyne, & Harn, 2003; Western Regional Reading First Technical Assistance Center, 2005). Lessons maps highlighted the most essential components of a lesson, provided time and grouping guidance for each component, and included recommendations for content to delete or deemphasize (Baker et al., 2010). Teachers received training and coaching support to incorporate the use of the lesson maps into their Tier 1 instruction. To help teachers with curriculum coverage, a pacing guide mapped out which lesson map the teacher should use on each day of Tier 1 reading instruction.

Second, instructional templates or teaching routines were used to enhance instruction quality, explicitness, and scaffolds (Baker et al., 2010; Western Regional Reading First Technical Assistance Center, 2005). The templates provided routines for teachers on how to teach specific skills. The 11 templates provided specific wording to introduce and practice regular and irregular word reading, phoneme segmentation and blending, spelling by sound, sound-by-sound blending, continuous blending, and word reading with an emphasis on spelling. Three of the templates also gave teachers specific routines for reading decodable text with students. Recommended routines were based on level of student performance.

Each template was organized the same way and included six features: (a) a summary of the materials needed for instruction, and the wording and signaling procedures that enhance instruction; (b) a clear explanation of the task teachers should provide to students prior to starting the activity; (c) details of how to model the task; (d) directions on how to lead student discussions and provide group practice using specific verbal and visual cues (e.g., "Your turn"); (e) directions on how to provide individual student practice; and (f) directions on how to correct student errors.

Tier 1 Professional Development. Tier 1 teachers participated in 5 days of professional development—3 days during the summer and 2 days in the fall. The 5 days addressed the science of beginning reading, effective instruction, and enhancing core, Tier 1 instruction through the use of ECRI instructional templates and lesson maps. Throughout the training teachers practiced using the templates and lesson maps with specific feedback.

Tier 1 Implementation Fidelity. ECRI staff conducted fidelity observations in Tier 1 settings. Observers were trained on the protocol for 5 days—2 days in the fall, 2 days in the winter, and 1 day in the spring. Observers were required to demonstrate at least 90% proficiency on the observation protocol prior to conducting independent observations in classrooms. In the field, observers and senior project staff conducted joint observations, and observers were required to achieve at least 85% reliability before collecting observation data that would be used in the analysis. Approximately 20% of all observations had two

Table 2. Examples of general ECRI enhancements of Tier 1 and Tier 2 instruction

ECRI Enhancements	Tier 1 Support	Tier 2 Support
1. State learning objectives clearly for students.	Teachers strategically explain the purpose of the lesson and what students will learn. "Now you're going to learn to blend individual sounds to make words."	Tier 2 learning objectives are the same as Tier 1. Instructors clearly state the skill students are learning, which the student was likely to have heard that day during Tier 1.
2. Increase the amount and quality of teacher models. Models include what teachers expect students to do through visual demonstrations, verbal directions and clear explanations.	Teachers are trained to strategically use specific scripts and modeling procedures. With this method struggling learners can focus on the content and not the activity directions. "When I touch under a letter you'll say the sounds for that letter. When you blend, don't stop between the sounds. When I slide under the whole word, you'll say the word. I'll model for you how to blend two words. My turn."	Instructors are provided with the same instructional routines being used in Tier 1. In Tier 2 students get to practice and the instructor provides the same models that students have seen in Tier 1. This approach minimizes confusion for struggling readers.
3. Make and review connections systematically between new and previously learned concepts.	Lesson maps provide teachers with the specific content that should be reviewed from previous weeks and units. Teachers stop at strategic points throughout the lesson to connect concepts for students.	Instructors are provided the same content that the Tier 1 teachers are introducing. Tier 2 instructors remind students of connections between different content that they are learning or that they have already learned.
4. Increase the amount of guided (teachers and students together) and independent student practice and thereby ensuring that high levels of student accuracy could be attained.	Lesson maps give teachers specific strategies to increase student practice. Teachers are taught to identify activities in the core program that are not required for student learning. For example, within a specific week in a given unit teachers are instructed to offer 15 min of structured partner reading to target reading fluency. An instructional routine is provided and the student materials (e.g., decodable text) are selected for the teachers.	Within each Tier 2 session, instructors have a set number of highly engaging preview and practice activities for students. On a typical day students will be guided through seven different activities with content based on the Tier 1 scope and sequence.
5. Increase the amount of previously learned material and strategically incorporating that material throughout the lesson.	Every day the lesson map includes specific time spent on reviewing previously taught material. One day of each week is spent almost entirely on reviewing content taught earlier that week and from previous units.	Tier 2 instruction consists of review content but instructors also provide pre-teaching opportunities, thus priming students for learning in Tier 1. If tomorrow's Tier 1 words are end in silent "e" the Tier 2 teachers will introduce the concept to students in Tier 2 the day before.

Note. ECRI = Enhanced Core Reading Instruction.

observers conducting concurrent observations. Reliability was greater than 90% on each observation.

During the observations, Tier 1 teachers were observed using the lesson maps and templates 87% and 88% of the time, respectively. Instruction on foundational skills was observed in the majority of lessons: alphabetic principle = 90% of lessons (*SD* = .16), word reading = 89% of lessons (*SD* = .16), and phonemic awareness = 81% of lessons (*SD* = .23). Instruction in reading connected text was observed in 79% of the lessons (*SD* = .23). Comprehension instruction was observed in 58% of lessons (*SD* = .32) and vocabulary instruction was observed in 55% of lessons (*SD* = .32). Tier 1 teachers were rated as implementing the intended instructional delivery enhancements with a *high* degree of fidelity in 82% of observations (*SD* = .17).

Tier 2 Intervention Overview

The ECRI Tier 2 intervention supplemented Tier 1 instruction by providing an additional 30 min of small group instruction daily. Two instructional objectives guided Tier 2 intervention. First, students reviewed and practiced content taught in Tier 1. Second, students received instruction in upcoming content (preview) that was going to be taught in Tier 1. In this manner, Tier 2 intervention was highly aligned with Tier 1 instruction. Also, the same instructional templates were used in Tier 1 instruction and Tier 2 intervention.

Tier 2 Instructional Delivery. In addition to instructional delivery (see Table 1), Tier 2 intervention strongly emphasized two additional features: (a) the pace of instruction was brisk, a goal necessitating that students respond accurately a high percentage of the time so that students were not spending a lot of time unsure how to respond and teachers were not spending a lot of time correcting student errors, and (b) transition time between activities was minimized to facilitate numerous activities during the 30-min lessons to maintain student engagement.

Tier 2 intervention also included mastery monitoring or what has also been referred to as dynamic assessment (Oslund et al., 2012; Watts-Taffe et al., 2012). Content mastery was checked throughout each lesson. Interventionists maintained a data log, noting specific content students struggled with. Each lesson ended with 3 min of review of challenging content interventionists noted in their data logs. In addition, a set of "quick checks" were administered to students at the end of each unit. Interventionists used these data to guide instruction. For example, if three of five students did not pass the mastery check the interventionist added distributed review of the missed content over the course of the following week.

Tier 2 Intervention Content. ECRI interventionists followed an explicit sequence of seven instructional routines. Next is an overview of each routine and the approximate number of minutes spent on each.

Irregular word reading (2 min). The lesson began with irregular word reading that included common phonetically irregular (nondecodable; e.g., "the") and phonetically regular words for which all the sound spellings had not yet been taught (decodable; e.g., "them" when /th/ had not yet been introduced).

Phonemic awareness (1 min). The lesson included eight words used with the phoneme blending routine and eight additional words used with the phoneme segmentation routine. Words used for blending and segmenting practice either were from the core program or were added because they contained new and previously taught sound spellings.

Sound spelling introduction and review (2 min). After phonemic awareness activities, the new sound spelling was practiced and previously taught sound spellings were reviewed. Sound-spelling cards from the core program were used.

Blending and word reading (4 min). Blending included eight words used with a sound-by-sound blending routine to practice decoding words that included new and previously taught sound spellings. Words used for blending practice were selected from core program texts or blending practice activities. Words practiced had been used for blending practice on a previous day.

Accuracy and fluency reading decodable text (12 min). Nearly half of the lesson focused on reading decodable text for the purposes of applying decoding skills and building accuracy and fluency in connected text. Decodable readers from the core program were used so that the decodable text corresponded to the phonic elements being introduced in the Tier 1 lesson. As students' reading fluency reached end-of-year goals, they transitioned to an advanced routine that provided less think time between words to enable students to read accurately and fluently in the text.

Encoding practice (5 min). Encoding practice involved explicit instruction to spell four to six words that included the sound spelling introduced in the lesson and previously taught sound spellings. The words for encoding practice were used in a previous day's lesson for blending and regular word reading practice.

Reteaching of challenging words (3 min). The final lesson activity involved practicing one to eight sound spellings and words that students had pronounced or read incorrectly during the small-group lesson.

Tier 2 Interventionist Training. Tier 2 interventionists were trained for 2 days in the fall before the intervention began. A follow-up half day of training was provided in the winter. Tier 2 interventionists also received coaching visits at least monthly to monitor fidelity and provide additional training to ensure implementation with fidelity at all times. Coaching activities included observations, feedback on the delivery of instruction, and modeling of instruction. After observations, expert coaches provided additional written feedback to interventionists and met with them individually to review the feedback.

Tier 2 Implementation Fidelity. ECRI expert coaches conducted regular fidelity observations to monitor implementation but fidelity observations were not conducted by independent observers. ECRI expert coaches provided regular feedback to the Tier 2 interventionists, who were paid ECRI staff members, to ensure implementation with fidelity. The data collected by the ECRI coaches were not collected for use in the analysis.

ECRI coaches reported to the principal investigators regularly on two Tier 2 implementation fidelity features: (a) basic implementation features (e.g., following all steps in the lesson) and (b) advanced implementation features (e.g., increasing pacing of the lesson to increase student engagement). Because interventionists met criteria on basic implementation features, ECRI coaches focused on working with them during their regular visits on advanced implementation features. Providing additional practice for students was a common advanced implementation topic. In this case, the coach would model pacing and interactive instruction to facilitate high rates of practice. Then, during the lesson with students, the interventionist would have an opportunity to demonstrate these techniques, with direct feedback from the coach. In meetings with the PIs, coaches reported that all interventionists maintained acceptable implementation of advanced features of instruction.

Measures

Stanford Achievement Test Tenth Edition (SAT10)

The SAT10 (Harcourt Educational Measurement, 2002) was used to determine tier assignment at pretest and as the outcome measure at posttest. The SAT10 is a group administered, norm-referenced test of overall reading proficiency, and in Grade 1 the SAT10 manual reports that the overall Kuder-Richardson reliability coefficient for total reading score was .97. The manual also reports that correlations between the total reading score and the Otis-Lennon School Ability Test ranged from .61 to .74. In this study, scaled scores were used in all analyses.

Dynamic Indicators of Basic Early Literacy Skills: Oral Reading Fluency (DIBELS ORF)

DIBELS ORF (Good & Kaminski, 2002) was used to measure students' fluency with reading connected text. Students were administered three passages at the end of the year. Students' median words read correct scores from the three passages were used for analysis ($M = 75.4$, $SD = 37.6$). On DIBELS ORF passages, reported alternate-form reliability drawn ranged from .89 to .94 and test–retest reliabilities for elementary students ranged from .92 to .97 (Good & Kaminski, 2002). Baker et al. (2008) reported concurrent validity coefficient of .82 between DIBELS ORF and the SAT10 for students in Grade 1, and .80 for students in Grade 2.

Data Collection

Project staff collected all data used in student data analysis and measuring implementation fidelity. Supervisory staff included one part-time research coordinator and three part-time assessment and observation coordinators. A total of 43 data collectors collected student reading data and implementation fidelity data.

Student Reading Data

Data collectors were trained to administer, score, and enter all student reading data used in the analysis. Training occurred during 2 days of initial training in the summer prior to beginning the study and 3 days of training during the study. Reading assessments were administered following standardized administration rules. Across the study, 20% of the reading assessments were observed by senior project staff, who also completed fidelity checklists. Average fidelity was 96%.

Analysis Methods

RD design and analysis uses a single score on a continuous assignment variable to determine the dichotomous treatment indicator. The "assignment variable may be any variable measured before treatment, including the pretest on the outcome variable" (Shadish, Cook, & Campbell, 2002, p. 209). The treatment indicator, in turn, governs whether children receive or do not receive the intervention. The approach then compares students with scores immediately above and immediately below the dichotomous treatment indicator while

controlling for the assignment variable to account for selection effects. A statistically significant difference would indicate that students who did not receive the Tier 2 intervention would have likely benefitted from the intervention had they received it. That is, those students with scores just below the treatment indicator benefited from the Tier 2 intervention over those who received only Tier 1. Unlike a randomized trial that compares the potential outcomes (Rubin, 1974) across two similar, randomly assigned groups of students, an RD study compares students just above with students just below the scores that determine the dichotomous treatment indicator.

The design and analysis of this RD study follow the recommendations in the methods literature (e.g., Bloom, 2012; Imbens & Lemieux, 2008; Judd & Kenny, 1981; Shadish et al., 2002). Experimental control with this design has been demonstrated by Rubin (1977), and Shadish et al. (2002) show "why this design ... capitalizes on selection but still provides unbiased causal inferences" (p. 207). Judd and Kenny (1981) and Rubin (1977) also show that the RD model avoids biases such as regression toward the mean, the unreliability of dichotomous treatment indicator, and unmeasured background variables.

The RD analysis depends heavily, however, on the assumption that we have correctly modeled the underlying relationship between the continuous assignment variable and the outcomes. Incorrectly modeling that relationship can lead to the appearance of a discontinuity (intervention effect) when none exists. If that relationship has been correctly modeled under the assumption of no Tier 1/Tier 2 discontinuity, then the treatment estimates would be unbiased (Rubin, 1977). We used schools randomly assigned to control, with respect to the Tier 2 intervention—unavailable in most RD studies—to test the relationship under the no-discontinuity assumption. This approach rules out the effects of any other type of instruction typical in Grade 1 classrooms, such as the default provision of Tier 2 instruction, the various Tier 3 interventions, or other forms of individual or differentiated instruction typical in the districts involved in this study. Hence, by modeling the relationship between the assignment variable and outcome in the control school sample, the size of the discontinuity, represented by the treatment indicator, is expected to provide an unbiased estimate of the difference between Tier 1 and Tier 2 for those students with scores surrounding the criterion for the treatment indicator given all existing, unrelated activities in the first-grade classrooms.

Specifically, the data analysis attempts to identify a statistically significant discontinuity at the cut point associated with assignment to the Tier 2 small-group intervention. The analysis relied on standard regression procedures. We regressed each dependent variable, Y, collected at posttest on the dichotomous treatment indicator, X, the continuous assignment variable, Z, and additional covariates, C.

$$\mathrm{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3(\mathrm{Z}{-}494)^2 + \beta_4 C + \varepsilon.$$

The SAT10 Total Reading scaled score administered at pretest was the assignment variable, Z, where students who scored at or below the 30th percentile were assigned to small-group intervention. Thus, $X = 1$ for cases with a Total Reading scaled score below 488 and $X = 0$ otherwise. The RD analysis also included Z squared, after subtracting the median (494) and, for dependent variables other than the Total Reading score that were collected in the fall; we also entered their pretest values as a covariate.

The RD analysis relies on a number of assumptions. First, the ECRI project used a sharp RD design such that all students assigned to Tier 2 received the small-group intervention. Second, the analysis included complete cases, students with both pretest and posttest data. Third, it is important to capture the correct functional form assignment variable when

regressed on the dependent variable. To ensure we used the correct form of the relationship between the assignment variable and the outcome, we estimated this relationship with the control schools. Within control schools, we were able to test the higher order polynomial functions (e.g., Z^2, Z^3) to identify the most predictive relationship between measures and avoid potential spurious effects.

RESULTS

Table 3 provides descriptive information and correlations between all measures. No parents withdrew their students from the study, but SAT10 scores, our primary outcome, were available for only 93% of students at posttest, 95% for those students above the cut score and 91% for those below. Although pretest scores clearly differed between those above and below the treatment indicator cut point, we conducted an analysis to test whether pretest scores were differentially affected by missingness across conditions. The results indicated that the difference was not moderated by condition, implying no differential effects of missing data by condition, for the SAT10 total score and all three subscales available at pretest.

Functional Relationship Between the Predictor and the Outcome Measure

Using the sample of students in control schools, we predicted the SAT10 Total Reading score in the spring with linear, quadratic, cubic, quartic, and higher order terms for the assignment variable, fall SAT10 Total Reading score. The model with only a linear term accounted for 57.5% (adjusted R^2) of the variation in the total score. The model with linear and quadratic terms improved the model fit, accounting for 59.1% of the posttest variation. Cubic and quartic terms improved the prediction but only marginally, increasing the adjusted R^2 from .5909 to .5912. In addition to the slight change in adjusted R^2, we also examined the AIC and BIC (Burnham & Anderson, 2002) and found negligible improvements for higher order terms beyond the quadratic, so we retained the simpler, two-term model with the SAT10 Total Reading score and its squared value for our RD analyses. The overall shape of the relationship was increasing with a slightly concave, decelerating trend.

Intervention Effects

The primary research question is whether students at risk for reading difficulty and receive Tier 1 instruction plus Tier 2 intervention outperform similar students who receive Tier 1 instruction only. We tested the effects on the SAT10 Total Reading score, SAT10 subtests, and oral reading fluency. The analyses produced statistically significant intervention effects—statistically significant regression discontinuities at the cut point defined by the treatment indicator—for the SAT10 Total Reading score ($p = .001$). There was a significant difference on each of the SAT10 subtests, Word Study ($p = .031$), Word Reading ($p = .049$), Sentence Reading ($p = .003$), and Reading Comprehension ($p = .022$), but we found no statistically significant difference for oral reading fluency. Table 4 displays the regression model results for each dependent measure.

The chances of a Type I error increases for multiple tests. We tested two independent measures, SAT10 and oral reading fluency, with four subtests of the SAT10. With six

Table 3. Means and standard deviations of the SAT10 total and subtest scores and oral reading fluency with correlations for all students

			Correlations									
Measure	*M*	*SD*	1	2	3	4	5	6	7	8	9	10
Fall assessments												
1. SAT10 Total	505.07	(57.43)	—									
2. SAT10 Sounds and Letters	533.02	(45.94)	.85	—								
3. SAT10 Word Reading	470.21	(57.27)	.94	.72	—							
4. SAT10 Sentence Reading	498.79	(63.07)	.93	.70	.86	—						
5. Oral Reading Fluency Words Correct	11.23	(8.77)	.42	.01	.38	.32	—					
Spring assessments												
6. SAT10 Total	564.03	(43.73)	.76	.70	.71	.72	.28	—				
7. SAT10 Word Study Skills	581.79	(50.50)	.66	.61	.61	.60	.13	.85	—			
8. SAT10 Word Reading	552.89	(51.83)	.67	.60	.65	.66	.30	.84	.62	—		
9. SAT10 Sentence Reading	564.38	(44.54)	.58	.57	.54	.56	.16	.82	.58	.67	—	
10. SAT10 Reading Comprehension	567.88	(49.16)	.71	.64	.67	.68	.28	.91	.67	.76	.71	—
11. Oral Reading Fluency Words Correct	75.44	(37.55)	.74	.54	.72	.70	.51	.73	.56	.67	.51	.72

Note. All correlations are statistically significant ($p < .0001$).

Table 4. Regression discontinuity model results for all dependent variables

		SAT10					
Variable		Total Reading	Word Study	Word Reading	Sentence Reading	Reading Comp.	ORF
Intercept		190.8 (16.7)	244.5 (23.1)	121.4 (22.0)	220.8 (21.6)	174.6 (20.1)	−208.7 (80.1)
Fall SAT10 total reading	Z	0.7422 (0.0319)	0.6652 (0.0441)	0.9548 (0.0652)	0.7719 (0.0616)	0.7833 (0.0385)	0.5574 (0.0409)
Fall SAT10 total reading squared	Z^2	−0.0023 (0.0002)	−0.0013 (0.0003)	−0.0045 (0.0003)	−0.0033 (0.0003)	−0.0025 (0.0003)	−0.0054 (0.0022)
DV collected at pretest[a]	C			−0.0814 (0.0515)	−0.0844 (0.0462)		
Treatment	X	9.77 (2.98)	8.88 (4.12)	7.76 (3.93)	11.98 (3.95)	8.27 (3.59)	3.72 (3.15)
Treatment *p* value		.001	.031	.049	.003	.022	.238
Adjusted R^2		.58	.40	.51	.36	.53	.53
Model *F* value		652.0	307.3	359.8	194.8	519.6	389.3

Note. All model *F* values were statistically significant with $p < .0001$. Fall SAT10 total reading and total reading squared was statistically significant at $p < .0001$ for all models except oral reading fluency (ORF). Pretest covariates were not statistically significant for word reading ($p = .1140$) or sentence reading ($p = .0681$). The SAT10 terms were not statistically significant predictors for the dependent variable (DV) ORF, but fall ORF was statistically significant ($p < .0001$). SAT10 models included 1,393 students; the ORF model included 1,144 students.

[a]The pretest covariate was the same as the dependent variable and included in the model only if it was available and not already in the model (i.e., Total Reading).

tests, we would expect a .27 chance of one or more errors (.73 chance of no errors), or between zero and two Type I errors with 95% confidence. Alternatively, assuming a strict, conservative Bonferroni correction for our *t*-test criterion ($\alpha < .008$), we would conclude that we found statistically significant results only for the SAT10 and its Sentence Reading subtest. We believe the pattern of results supports the conclusion that students who received Tiers 1 and 2 improved on all SAT10 measures relative to those who only received Tier 1.

Figure 1 shows the results for the Total Reading score. It depicts the observed values and the lines predicted by the regression analysis, which as shown in Table 4, differs by 9.77 points. That is, students receiving both Tier 1 instruction and Tier 2 intervention achieved significantly higher reading scores at the cut point compared to students who received Tier 1 only. In practical terms, the difference is approximately 8 percentile points. At the cut score, students in the control group would have earned a score at approximately the 31st percentile and students in the treatment group would have scored at the 39th percentile.

Sensitivity Analysis of Intervention Effects

We followed this by conducting a sensitivity analysis within the treatment sample, modeling the intervention effects with and without the cubic and quartic terms. These were the higher order terms that improved model fit, albeit only slightly. The presence of these higher order terms did not appreciably change the effect of the treatment indicator. Similar models were fit for each dependent variable with similar results, and the higher order terms did not influence the conclusions for any of them.

We also tested whether there was a break in the regression line in the control schools, expecting that intervention effects should be specific to intervention schools and not appear in control schools. Control schools were treated the same as intervention schools except that they provided the Tier 2 supports that the district and school typically offered. To test whether control students in Tier 2 received a treatment effect, we tested for intervention effects within the control classrooms for all dependent measures using the same quadratic prediction model between the dependent variable and assignment variable. None of the analyses produced statistically significant discontinuities associated with the treatment indicator. Figure 2 shows the results of this analysis for the SAT10 Total Reading score.

Finally, we tested the heterogeneity of the treatment effect through a series of multilevel models that nested the intercept and treatment indicators, X, within classrooms using SAS PROC MIXED (SAS Institute, Inc., 2009). Heterogeneity of the treatment effect is represented by a statistically significant estimate of the variance for treatment between classrooms. We found no evidence of treatment heterogeneity for any the outcomes by classroom. We repeated the analysis with schools as the level of clustering and again found no evidence of treatment heterogeneity by school.

Note that although nesting treatment within classroom or school may indictate treatment variability, it should not affect the esimate or statistical significance of the fixed effect for treatement (Raudenbush & Sadoff, 2008). Although we observed small changes in the estimates, likely due to different estimation methods, we found no substantive changes in the results with one exception. For SAT10 Word Reading, the effect decreased and the p value increased in the multilevel models, from 7.76 ($p = .049$) estimated from the standard regression model to 6.70 ($p = .091$) in the multilevel model with classroom as a clustering variable, and 6.58 ($p = .105$) with school as the cluster.

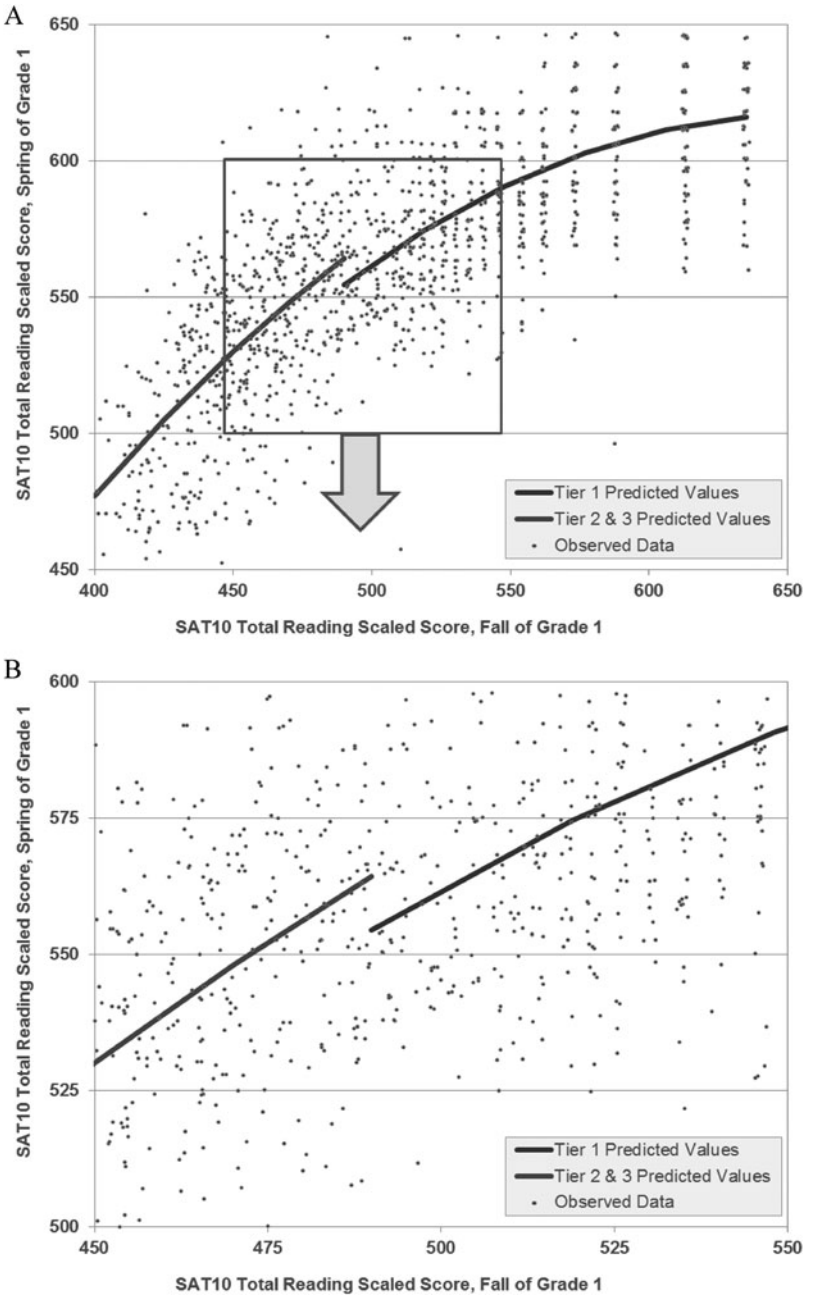

Figure 1. Observed values and predicted lines for the SAT10 total score regression discontinuity. *Note.* Plot A shows the full sample of students. Plot B shows a close-up view of the subset of values around the discontinuity—the sample of students within the rectangle in Plot A.

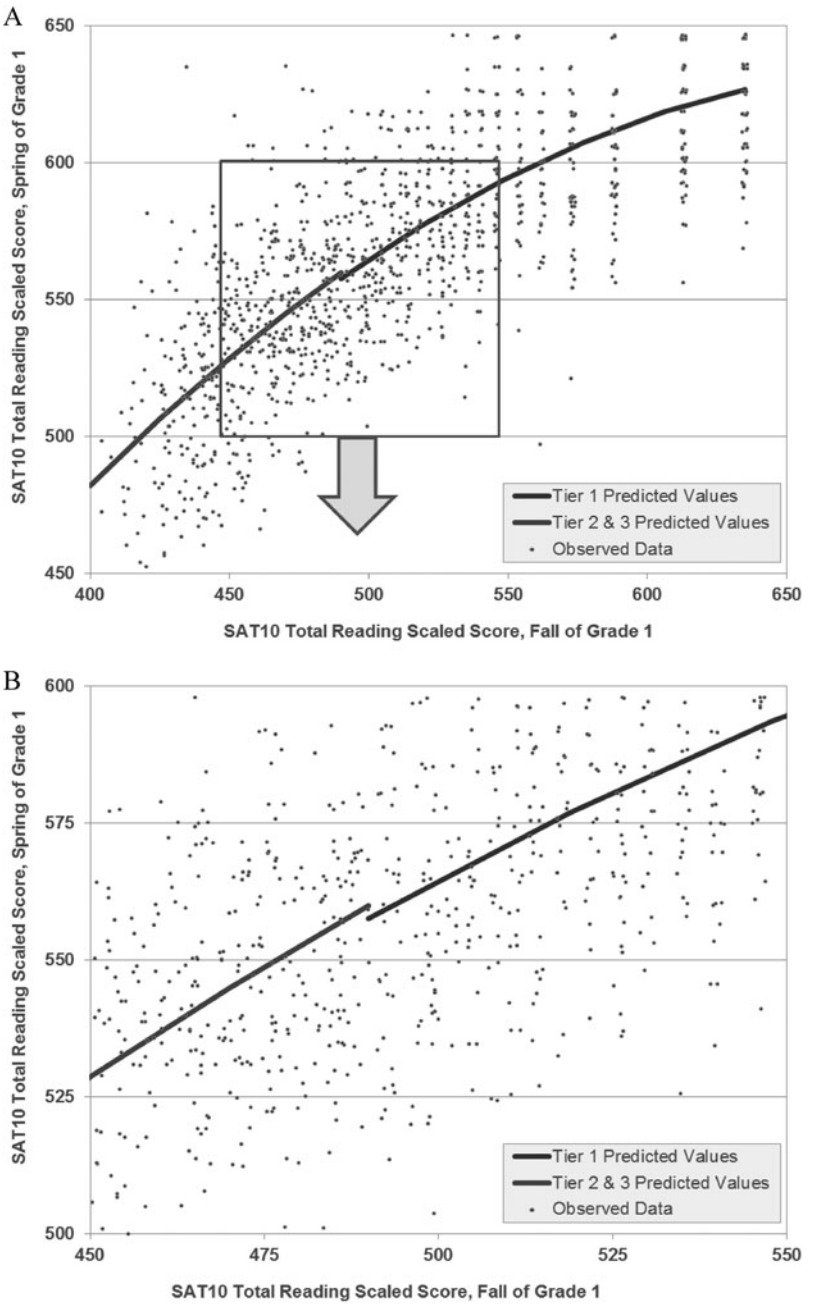

Figure 2. Observed values and predicted lines for SAT10 total score from regression discontinuity model for the ***control*** sample. *Note.* Plot A shows the full sample of students. Plot B shows a close-up view of the subset of values around the discontinuity—the sample of students within the rectangle in Plot A.

DISCUSSION

The purpose of this study was to test the impact of a Tier 2 reading intervention with students at risk for reading difficulties in first grade who just make or just miss the cut-off score to receive Tier 2 set at the 30th percentile on the pretest reading assessment. The major finding was that students who received Tier 2 made greater gains than did students who scored just above the cutoff score on the reading pretest and hence received Tier 1 instruction only. In the context of the RD design, the findings apply only to students close to but on either side of the cut score (Smolkowski et al., 2013). An assumption of the RD design is that students close to the cut score in the Tier 1 only (control) condition would have made the same gains as students close to the cut score in the treatment condition if they had received the intervention. In practical terms, receiving Tier 2 as well as Tier 1 boosted performance by approximately 8 percentile points. Receiving only Tier 1, students near the cut score would have been expected to be reading at approximately the 31st percentile on the SAT10 (about where they began so expected gains). With the addition of Tier 2, students near the cut score would be expected to be reading at approximately the 39th percentile (accelerated gains).

Findings in the Context of RTI

In the first ECRI study (Fien et al., 2014; Smith et al., 2013), students at risk for reading difficulties who received ECRI Tier 1 instruction with Tier 2 intervention outperformed similar students in a business-as-usual Tier 1 plus Tier 2 approach. This finding is consistent with previous Tier 2 intervention studies (Chambers et al., 2011; Gersten et al., 2009; O'Connor, Fulmer, Harty, & Bell, 2005; Vadasy et al., 2005). This study adds to previous findings by demonstrating that intervention students who received ECRI Tier 1 instruction and Tier 2 intervention outperform similar students, who missed the cutoff, and thus received ECRI Tier 1 instruction only. These students, instead of receiving a contrasting Tier 2 intervention (i.e., business as usual), received what they would typically receive in school settings, which is core, Tier 1 instruction. Thus, this study answers the question, what happens to students who are similarly skilled when some students who meet criteria (fall just below the cut point) for a Tier 2 intervention receive the intervention and students who do not meet the criteria (fall just above the cut point) do not receive the intervention? By employing an RD design, and because adherence to treatment and control conditions in this study was high (i.e., we were able to use a sharp design), we could address this question precisely. That is, by assigning students to the intervention condition and thereby making the determination these students were at risk for reading difficulties, the students in the Tier 1 only condition close to the cut score would also be at the same level of risk for reading difficulties.

The context in which students who are similarly skilled and have the same level of risk of reading difficulty, and yet some are placed into an intervention and others are not, points to a potential concern with RTI as a framework for decision making. That is, by using some kind of cut score to determine which students receive services, such as supplemental Tier 2 intervention, and which students do not, services would not be provided to students who may actually need them. In this study, for example, students in the control condition close to the cut score are from the same population as students in the treatment condition close to the cut score, with the same level of risk of reading difficulty and need for service. To conduct a valid RD study it was necessary to adhere to a rigid cut score to eliminate selection bias

(Rubin, 1977). Potentially, any efficacious intervention in an RD study produces differences at the chosen cut score.

In practice settings, however, schools are not concerned in the same way about selection bias and their decision making might increase the likelihood that RTI practices in school contexts are as responsive as possible to student need, although this has not been tested. For example, decisions about services might be made on the basis of multiple sources of information, including test scores and professional judgment. Thus, current level of performance, slope of change over time, and responsiveness to classroom instruction might all be considered in making RTI decisions about tiers of support. Real-time decisions about services can also be made on the basis on student responsiveness to instruction and intervention. Services do not have to be allocated on a fixed time basis as they are in the case of a sharp RD design. For instance, students in Tier 1 who are not making progress can receive Tier 2 services when the school team determines it is necessary. In this RD study, it was necessary for students to complete their participation in their assigned condition for the entire length of the intervention.

Another important consideration in terms of RTI practice in school settings is that the vast majority of students at risk for reading difficulty in the early grades who receive Tier 2 intervention also receive Tier 1 instruction alongside their peers who receive Tier 1 instruction only (Hill et al., 2012; Wanzek & Cavanaugh, 2012). Thus, our study provides potentially important information about school practices that are in widespread use. That is, schools that provide students at risk for reading difficulties with Tier 1 instruction and Tier 2 intervention, and expect the combination of services to have a positive impact, have an additional reason for optimism, especially if their practices are in line with what occurred in ECRI schools.

The short-term goal is that students at risk of reading difficulty receive a supplemental, Tier 2 intervention that accelerates their literacy growth. The long-term goal is that students catch up to their peers who are not receiving extra support and then are able to make progress that is on par with their peers when the support is removed. In this study, the addition of the Tier 2 intervention increased the reading gains of students in Tier 2 compared to their peers who received Tier 1 only, so that by the end of the intervention it is possible some of these students would no longer require Tier 2 intervention to make adequate reading growth. However, we do not report second-grade outcomes for these students and can only speculate about the reading services these students needed or received after the intervention.

This study addresses another important question in terms of how tiered systems are conceptualized and implemented in authentic school settings.

Evidence of Tier 2 Impact

In terms of tiered approaches to instruction and intervention, we suggest schools should consider evidence involving two types of outcomes as they make decisions about adopting a multitiered system or program. First, there should be evidence the system overall is efficacious compared to other approaches including business-as-usual practice. Second, in considering interventions for students at risk for reading difficulties, schools should examine the evidence that the intervention helps close the achievement gap with students who not at risk of reading difficulties.

Demonstrating these two types of outcomes is important to establishing the empirical evidence for tiered systems of instruction and intervention. Evidence in one of these two areas but not the other could represent a potential problem with the approach when

implemented by schools in authentic settings. In the context of Tier 2 interventions, for example, if the approach is no better than typical practice (i.e., compared to other Tier 2 interventions), then there is no compelling reason to adopt the intervention. In addition, if the Tier 2 intervention does not help students at risk for reading difficulties catch up to their peers who receive Tier 1 only, then the gap in reading performance will persist, which may present problems for those students in the classroom if the extra support is removed (Rodden-Nord et al., 1992; Shinn, Baker, et al., 1993; Shinn, Habedank, et al., 1993).

The Institute of Education Sciences Practice Guide identifies two types of interventions for Tier 2 students. The first is to provide a supplemental Tier 2 intervention for students, which was the focus of this study. The second is to provide differentiated instruction in Tier 1 based on the type of instruction students need. For example, Connor and colleagues (2009) examined the impact of differentiated instruction (referred to as individualized) and observed that students in classrooms who received specified amounts of differentiated instruction may have outperformed students in classrooms who received less precise, less differentiated instruction. The ECRI intervention included enhancing Tier 1 instruction to make it more accessible to students at risk for reading difficulties and Tier 1 teachers were trained to differentiate instruction for students based on need, but we did not manipulate differentiated instruction in Tier 1 to test its effects. Future studies should examine the value of differentiated instruction in Tier 1 in relation to Tier 2 interventions, as well as the effects of integrating differentiated instruction in Tier 1 with Tier 2 interventions.

Using Regression Discontinuity Designs

In examining the impact of tiered systems of instruction and intervention, RD and RCTs have complimentary roles to play. RCTs can help determine the impact of innovative interventions versus BAU (e.g., Tier 1 vs. Tier 1; Tier 2 vs. Tier 2) approaches. RD can help determine if the increased intensity of instruction and intervention in the higher tier (e.g., Tier 2 vs. Tier 1) accelerates growth among students.

These designs can work in the context of how schools operate, although in the absence of resources and the personnel necessary to design and conduct these studies correctly the assumptions underlying their use can be easily compromised. For example, districts regularly adopt new programs and materials, often as every 5 to 7 years, for widespread implementation in schools throughout the district. In a district with 30 elementary schools, instead of having all schools begin implementation of the new program in the 1st year or staggering implementation by year based on school choice, the start year could be determined randomly with half of the schools beginning the program in Year 1 and the remaining half of the schools beginning the program in Year 2.

Within a school, RD could be used to evaluate whether students who receive more intense instruction and intervention improve relative to comparable peers, similar to the way the current study was conducted. Given the increase in the use of tiered instruction and intervention approaches among districts and schools throughout the country, it would seem RD would be a useful approach to evaluate intervention programs within a tiered system. RD evaluations can even accommodate assignment decisions based on multiple scores (Reardon & Robinson, 2012).

Strict adherence to RD design assumptions requires that students stay in their assigned tier or tiered combination for the duration of the investigation if the objective is to assign causal attribution to intervention impact. This requirement will be difficult for many schools, primarily because it conflicts with principles they hold about regrouping students

for instruction and intervention based on what they believe students need, sometimes immediately. For example, school personnel feel compelled to make what they view as timely decisions for students who are not making adequate growth, by providing them with more intense instruction, including moving them to a more intense tier, as soon as the students need it. In fact, there may be evidence to suggest that the sooner students are provided with more intense service the greater their long-term gains (Al Otaiba et al., 2011; Al Otaiba et al. 2014).

Simultaneously—although we suspect this is less frequently the case—schools feel compelled to move students to less intense tiers of instruction and intervention when they believe students no longer need the service. The rationale frequently has to do with the equitable and responsible use of increasingly scarce and expensive resources. Of course, disruptions in group composition prior to the completion of a study is not only a problem for RD, it is also a problem for RCTs and quasi-experiments. Solutions must balance the questions asked, the designs used to answer these questions, and practical considerations underlying the contexts in which the investigations are being conducted. For example, in this study students remained in their assigned condition for the year-long intervention. In a different context, the assignment period could be much less than 1 year.

Limitations

Several limitations are important in considering the findings in this study. The districts and schools selected were Title I schools with many features of multitiered instructional systems in place as a condition of recruitment for the study. Although an increasing number of schools are moving to tiered systems of instruction and intervention, the findings for schools that do not use a tiered system may not be as relevant.

The intervention included a system of components, and we do not know which of the components contributed to the outcomes. For example, the intervention included manipulations to align Tier 1 instruction and Tier 2 intervention to make instruction and intervention more accessible to students at risk for reading difficulty. It is not clear if the same outcomes would have occurred if the alignment of Tier 1 and Tier 2 had not been a priority.

Many multitiered systems of instruction and intervention are more fluid in terms of their placement of students into interventions than we could allow in order to meet the assumptions of RD. The argument for more dynamic decision making is persuasive (Al Otaiba et al., 2011; Compton et al., 2010), but the case for making substantive changes in tiered placements carefully and deliberately, and on the basis of compelling data (including fidelity of implementation data), also has merit. It is not clear how a more flexible system of moving students into and out of specific tiers would influence the outcomes we observed in this study.

A final limitation concerns the measure used to place students into tiers, the SAT10. We chose a highly precise measure to determine placement because we felt the precision it offered to address the research question was important. The SAT10 is too lengthy and expensive for most schools to use as a screening measure. Schools typically use a more efficient and inexpensive test, or test battery, to screen first-grade students for reading problems. Other measures for screening students for reading problems might identify other students Tier 1 instruction versus Tier 1 plus Tier 2 intervention than were identified in this study, which could lead to different study outcomes. We do not think the use of different psychometrically sound screening instruments would identify substantively different students for Tier 2, although the placement of the cut score would result in more

or less students being identified. In particular, replicating this study using different measures and approaches for screening students into tiers, and perhaps using different cut scores to determine risk, are important areas for additional research.

Conclusion

To the best of our knowledge, this is the first RTI study to examine whether students who receive Tier 1 instruction and a Tier 2 intervention gain more in terms of reading achievement than students who receive Tier 1 instruction only. This outcome may have occurred in previous Tier 2 intervention studies in which students in a Tier 2 treatment condition outperformed control students in Tier 2, but these studies have not investigated the performance of students in Tier 2 relative to students in Tier 1.

We propose two explanations for the greater gains made by students in the treatment condition in this study. The first is time on task (Connor et al., 2009; Greenwood, 1991). Treatment students received the same Tier 1 instruction as control students plus an additional 30 min of Tier 2 intervention. The extra time in reading instruction may have accelerated the reading achievement of treatment students. However, many interventions include additional time yet do not yield an improvement in student literacy outcomes.

A second plausible explanation is that the alignment in content and delivery between Tiers 1 and 2 benefitted treatment students. Briefly, this explanation asserts that an important aspect of instruction for students at risk for reading difficulties is to increase academic engagement within a given amount of time (Greenwood, 1991). These students may not learn all the material taught during Tier 1 instruction because the intensity of instruction is not sufficient for complete or deep processing (Perfetti, 2007; Warren, Fey, & Yoder, 2007). Supplemental Tier 2 intervention offers an opportunity for more intense instruction. It is important that the content in Tier 2 be essentially the same content presented in Tier 1, given the hypothesis that the target students did not learn this material in a deep way during Tier 1 instruction. The extra time and more intense format (i.e., small-group intervention) provides students with additional opportunities to learn critical content and practice applying it in familiar and new learning situations.

It is also plausible that there is an interaction between the additional time in Tier 2 intervention and the alignment between instructional content and delivery in Tiers 1 and 2. In the context of this proposed interaction, we would predict that increasing both time in Tier 2 intervention and the content and delivery alignment between tiers would result in the greatest gains in reading growth over the course of the intervention. Experiments can help determine the legitimacy of these predictions. We hope future studies will explore issues related to engaged time and content alignment between tiers, and the possibility these variables might interact in some synergistic way to produce gains that are greater than the sum of the individual components.

FUNDING

This research was supported by Grant R324A090104 from the Institute of Education Sciences, U.S. Department of Education to the University of Oregon. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

REFERENCES

Al Otaiba, S., Connor, C. M., Folsom, J. S., Greulich, L., Meadows, J., & Li, Z. (2011). Assessment data–informed guidance to individualize kindergarten reading instruction: Findings from a cluster-randomized control field trial. *The Elementary School Journal*, *111*, 535–560.

Al Otaiba, S., Connor, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K. (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children*, *81*, 11–27. doi:10.1177/0014402914532234

Ashlock Consulting. (2006). *Lesson maps*. Petaluma, CA: Author.

Baker, S. K., Fien, H., & Baker, D. L. (2010). Robust reading instruction in the early grades: Conceptual and practical issues in the integration and evaluation of Tier 1 and Tier 2 instructional supports. *Focus on Exceptional Children*, *42*, 1–20.

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J., Kame'enui, E., & Thomas Beck, C. (2008). Reading fluency as a predictor of reading proficiency in low-performing high poverty schools. *School Psychology Review*, *37*, 18–37.

Bloom, H. S. (2012). Modern regression discontinuity analysis. *Journal of Research on Educational Effectiveness*, *5*, 43–82. doi:10.1080/19345747.2011.578707

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach (*2nd *ed.)*. New York, NY: Springer-Verlag.

Burns, M. K., & Gibbons, K. A. (2012). *Implementing Response-to-Intervention in elementary and secondary schools: Procedures to assure scientific-based practices* (2nd ed.). New York, NY: Routledge.

Chambers, B., Slavin, R. E., Madden, N. A., Abrami, P., Logan, M. K., & Gifford, R. (2011). Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second graders. *The Elementary School Journal*, *111*, 625–640. doi:10.1086/659035

Chard, D. J., Harn, B., Sugai, G., Horner, R., Simmons, D. C., & Kame'enui, E. J. (2008). Core features of multi-tiered systems of reading and behavioral support. In C. R. Greenwood, T. R. Kratochwill, & M. Clements (Eds.), *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 31–58). New York, NY: Guilford.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., . . . Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, *102*, 327–340. doi:10.1037/a0018448

Connor, C. M., Morrison, F. J., Fishman, B., Giuliani, S., Luck, M., Underwood, P. S., . . . Schatschneider, C. (2011). Testing the impact of child characteristics × instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*, 189–221. doi:10.2307/41228651

Connor, C. M., Piasta, S. B., Fishman, B., Glasney, S., Schatschneider, C., Crowe, E., . . . Morrison, F. J. (2009). Individualizing student instruction precisely: Effects of Child × Instruction interactions on first graders' literacy development. *Child Development*, *80*, 77–100. doi:10.1111/j.1467-8624.2008.01247.x

Coyne, M. D., Kame'enui, E. J., & Carnine, D. W. (2011). *Effective teaching strategies that accommodate diverse learners*. Upper Saddle River, NJ: Merrill.

Denton, C. A. (2012). Response to Intervention for reading difficulties in the primary grades. *Journal of Learning Disabilities*, *45*, 232–243. doi:10.1177/0022219412442155

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children*, *65*, 399–415.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*, 605–619. doi:10.1037/0022-0663.92.4.605

Evans, M. A. (1996). Reticent primary grade children and their more talkative peers: Verbal, nonverbal, and self-concept characteristics. *Journal of Educational Psychology*, *88*, 739–749.

Fien, H., Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson-Walker, N. J., & Chaparro, E. A. (2014). An examination of the efficacy of a multi-tiered intervention on early reading outcomes for first grade students at risk for reading difficulties. *Journal of Learning Disabilities.* Advance online publication. doi: 10.1177/0022219414521664

Fuchs, D., Fuchs, L. S., & Stecker, P. M. (2010). The "blurring" of special education in a new continuum of general education placements and services. *Exceptional Children*, *76*, 301–323.

Gersten, R. M., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2009). *Assisting students struggling with reading: Response to Intervention and multi-tier intervention for reading in the primary grades: A practice guide (*No. *NCE 2009-4045)*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, US Department of Education. Retrieved from http:/ies.ed.gov/ncee/wwc/publications/practiceguides/

Good, R. H., & Kaminski, R. (2002). *Dynamic indicators of basic early literacy skills (6*th *ed.)*. Eugene, OR: Institute for the Development of Education Achievement.

Greenwood, C. R. (1991). Longitudinal analysis of time, engagement, and achievement in at-risk versus non-risk students. *Exceptional Children*, *57*, 521–535.

Gunn, B., Smolkowski, K., Biglan, A., Black, C., & Blair, J. (2005). Fostering the development of reading skill through supplemental instruction: Results for Hispanic and non-Hispanic students. *Journal of Special Education*, *39*, 66–85. doi:10.1177/00224669050390020301

Gunn, B., Smolkowski, K., & Vadasy, P. (2011). Evaluating the effectiveness of Read Well kindergarten. *Journal of Research on Educational Effectiveness*, *4*, 53–86. doi:10.1080/19345747.2010.488716

Haager, D., Klingner, J. K., & Vaughn, S. (2007a). *Evidence-based reading practices for response to intervention*. Baltimore, MD: Brookes.

Haager, D., Klingner, J. K., & Vaughn, S. (2007b). *Validated reading practices for three tiers of intervention*. Baltimore, MD: Brookes.

Harcourt Educational Measurement. (2002). *Stanford Achievement Test [SAT-10]*. San Antonio, TX: Author.

Harn, B., Kame'enui, E. J., & Simmons, D. (2007). The nature and role of the third tier in a prevention model for kindergarten students. In D. Haager, S. Vaughn, & J. Klingner (Eds.), *Evidence-based reading practices for response to intervention* (pp. 161–184). New York, NY: Brookes.

Hill, D. R., King, S. A., Lemons, C. J., & Partanen, J. N. (2012). Fidelity of implementation and instructional alignment in response to intervention research. *Learning Disabilities Research & Practice*, *27*, 116–124. doi:10.1111/j.1540-5826.2012.00357.x

IDEA. (2004). Individuals with Disabilities Education Act of 2004, Public Law 108-446 Subpart 614(6)(b).

Imbens, G., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*, 615–635.

Judd, C. M., & Kenny, D. A. (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, *5*, 602–619.

Kame'enui, E. J., Simmons, D., Coyne, M., & Harn, B. (2003). *Institute on beginning reading: Day 1*. Eugene, OR: IDEA.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*, 75–86. doi:10.1207/s15326985ep4102_1

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J. L., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly*, *40*, 148–182. doi:10.1598/rrq.40.2.2

McMaster, K. L., Kung, S.-H., Han, I., & Cao, M. (2008). Peer-assisted learning strategies: A "Tier 1" approach promoting English Learners' response to intervention. *Exceptional Children*, *74*, 194–214.

National Association of State Directors of Special Education. (2006). *Response to Intervention: Policy considerations and implementation*. Alexandria, VA: Author.

O'Connor, R. E., Fulmer, D., Harty, K. R., & Bell, K. M. (2005). Layers of reading intervention in kindergarten through third grade: Changes in teaching and student outcomes. *Journal of Learning Disabilities*, *38*, 440–455.

Oslund, E. L., Hagan-Burke, S., Taylor, A. B., Simmons, D. C., Simmons, L., Kwok, O.-M., ... Coyne, M. D. (2012). Predicting kindergarteners' response to early reading intervention: an examination of progress-monitoring measures. *Reading Psychology*, *33*, 78–103. doi:10.1080/02702711.2012.630611

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*, 357–383. doi:10.1080/10888430701530730

Raudenbush, S. W., & Sadoff, S. (2008). Statistical inference when classroom quality is measured with error. *Journal of Research on Educational Effectiveness*, *1*, 138–154. doi:10.1080/19345740801982104

Reardon, S. F., & Robinson, J. P. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*, *5*, 83–104.

Rodden-Nord, K., Shinn, M. R., & Good, R. H. (1992). Effects of classroom performance data on general education teachers' willingness to reintegrate mildly handicapped students. *School Psychology Review*, *21*, 138–154.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*, 688–701. doi:10.1037/h0037350

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, *2*, 1–26. doi:10.2307/1164933

SAS Institute Inc. (2009). *SAS/STAT 9.2 user's guide*. Cary, NC: Author. Accessed from SAS Product Documentation website: http://support.sas.com/documentation/index.htmldoi:

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton-Mifflin.

Shinn, M. R., Baker, S. K., Habedank, L., & Good, R. H. (1993). The effects of classroom reading performance data on general education teachers' and parents' attitudes about reintegration. *Exceptionality*, *4*, 205–228. doi:10.1207/s15327035ex0404_1

Shinn, M. R., Habedank, L., & Baker, S. K. (1993). Reintegration as part of a problem-solving delivery service. *Exceptionality*, *4*, 245–251. doi:10.1207/s15327035ex0404_3

Smith, J. L. M., Fien, H., Nelson-Walker, N. J., Smolkowski, K., & Baker, S. K. (2013). *Examining the impact of a multi-tiered reading intervention in first grade: A clustered-randomized controlled trial*. Manuscript in preparation.

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student-Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly*, *27*, 316–328. doi:10.1016/j.ecresq.2011.09.004

Smolkowski, K., Strycker, L., & Seeley, J. R. (2013). The role of research evaluation for interventions on school-related behavior disorders. In H. M. Walker & F. M. Gresham (Eds.), *Handbook of evidence-based practices for students having emotional and behavioral disorders* (pp. 552–566). New York: Guilford.

Spectrum K12 School Solutions. (2010). *Response to Intervention adoption survey 2010*. Towson, MD: Author.

Sutherland, K. S., Alder, N., & Gunter, P. L. (2003). The effect of varying rates of opportunities to respond to academic requests on the classroom behavior of students with EBD. *Journal of Emotional and Behavioral Disorders*, *11*, 239–248. doi:10.1177/10634266030110040501

Taylor, B. M., Pearson, P. D., Clark, K. F., & Walpole, S. (1999). *Beating the odds in teaching all children to read* (Report No. 2-006). Ann Arbor, MI: Center for the Improvement of

Early Reading Achievement, University of Michigan. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED436723

Thurlow, M. L., Ysseldyke, J. E., Wotruba, J. W., & Algozzine, B. (1993). Instruction in special education classrooms under varying student-teacher ratios. *The Elementary School Journal*, *93*, 305–320. doi:10.2307/1001897

Vadasy, P. F., Sanders, E. A., & Peyton, J. A. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring. *Journal of Learning Disabilities*, *38*, 364–380. doi:10.1177/00222194050380041401

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities *Exceptional Children*, *69*, 391–411.

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Pedrotty Bryant, D., Dickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education*, *24*, 301–315. doi:10.1177/07419325030240050501

Vellutino, F., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Intervention as a vehicle for distinguishing between cognitive and experimental deficits as basic cause of specific reading disability. *Journal of Educational Psychology*, *88*, 601–638. doi:10.1037/0022-0663.88.4.601

Walker, H. M., & Shinn, M. R. (2002). Structuring school-based interventions to achieve integrated primary, secondary, and tertiary prevention goals for safe and effective schools. In H. M. Walker & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 1–26). Washington, DC: National Association of School Psychologists.

Wanzek, J., & Cavanaugh, C. (2012). Characteristics of general education reading interventions implemented in elementary schools for students with reading difficulties. *Remedial and Special Education*, *33*, 192–202. doi:10.1177/0741932510383162

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 70–77. doi:10.1002/mrdd.20139

Watts-Taffe, S., Laster, B. P., Broach, L., Marinak, B., McDonald Connor, C., & Walker-Dalhouse, D. (2012). Differentiated instruction: Making informed teacher decisions. *The Reading Teacher*, *66*, 303–314. doi:10.1002/trtr.01126

Western Regional Reading First Technical Assistance Center. (2005). *Instructional routine templates 1-17*. Eugene, OR: Author.

Wing, C., & Cook, T. D. (2013). Strengthening the regression discontinuity design using additional design elements: A within-study comparison. *Journal of Policy Analysis and Management*, *32*, 853–877.